

# PATENT ABSTRACTS OF JAPAN

BA

(11)Publication number : 2000-330979  
 (43)Date of publication of application : 30.11.2000

(51)Int.Cl. G06F 17/21  
 G06F 17/30

(21)Application number : 11-137090

(71)Applicant : NTT DATA CORP

(22)Date of filing : 18.05.1999

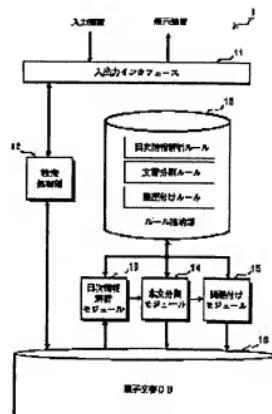
(72)Inventor : YAGI KIYOAKI  
 HARA KEIKO  
 KOMURO HIDEKAZU  
 YAMANAKA KOUSHI

## (54) METHOD FOR ANALYZING ELECTRONIC DOCUMENT TO BE RETRIEVED AND ELECTRONIC DOCUMENT REGISTRATION SYSTEM

### (57)Abstract:

**PROBLEM TO BE SOLVED:** To obtain the electronic document registration system which can register a document in such a style that a user can speedily reach desirable information.

**SOLUTION:** The system is equipped with an index information analysis module 13 which extracts index information included in a document file to analyze the structure of index items and give index levels based upon the structure analysis result to each index item, a body division module 14 which divides the body of the electronic file by the index levels according to the index levels and registers the divided body data as retrieval units, and a correlation module 15 which correlates individual key words included in a single piece of divided body data to the body data corresponding to the index item including the key word. Then the document file is divided into specific meaning contents units before being retrieved.



(19) 日本国特許庁 (JP)

## (12) 公開特許公報 (A)

(11) 特許出願公開番号

特許2000-330979

(P2000-330979A)

(43) 公開日 平成12年11月30日 (2000.11.30)

(51) Int.Cl?  
G O 6 F 17/21  
17/30

識別記号

F I  
G 0 6 F 15/20  
15/40マーク\* (参考)  
5 7 0 N 5 B 0 0 9  
5 8 0 J 5 B 0 7 5  
5 9 0 E  
3 4 0  
3 7 0 A

審査請求 未請求 請求項の数10 O.L. (全10頁)

(21) 出願番号 特願平11-137090

(71) 出願人 000102728

(22) 出願日 平成11年5月18日 (1999.5.18)

株式会社エヌ・ティ・ティ・データ  
東京都江東区豊洲三丁目3番3号

(72) 発明者 八木 清光

東京都江東区豊洲三丁目3番3号 株式会  
社エヌ・ティ・ティ・データ内

(73) 発明者 原 恵子

東京都江東区豊洲三丁目3番3号 株式会  
社エヌ・ティ・ティ・データ内

(74) 代理人 100099324

弁理士 鈴木 正剛

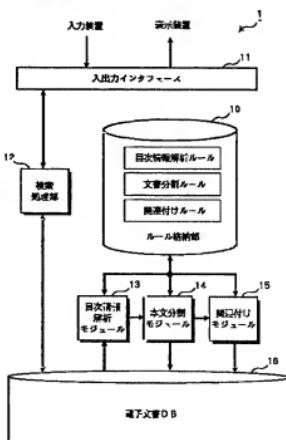
最終頁に続く

## (54) 【発明の名称】 検索対象となる電子文書の解析方法及び電子文書登録システム

## (57) 【要約】

【課題】 利用者が欲する情報に迅速に到達できるよう  
な形態で文書登録を行うことができる電子文書登録シス  
テムを提供する。

【解決手段】 文書ファイルに含まれる目次情報を抽出  
して目次項目の構造を解析し、各目次項目に構造解析結  
果に基づく目次レベルを付与する目次情報解析モジュール  
13と、目次レベルに基づいて文書ファイルの本文を  
目次レベル毎に分割し、分割後の本文データを検索単位  
として登録する本文分割モジュール14と、分割された  
本文データに含まれる個々のキーワードと当該キーワード  
を含む目次項目に対応する本文データとを相互に  
関連付ける関連付けモジュール15とを有し、文書ファ  
イルを検索前に所定の意味内容単位に分割しておくよう  
に構成する。



【特許請求の範囲】

【請求項1】 情報検索の対象となる電子文書から目次情報を抽出して当該電子文書における目次項目の構造解析を行い、各目次項目に前記構造解析結果に基づく目次レベルを付与するとともに、当該電子文書の本文を前記目次レベルに基づいて分割し、分割後の本文データを検索単位として登録することを特徴とする、電子文書の解析方法。

【請求項2】 前記目次項目が階層的な構造を有するものであり、前記目次レベルは、当該目次項目の階層関係を識別するための識別情報であることを特徴とする、請求項1記載の解析方法。

【請求項3】 前記電子文書は、前記目次項目が略定型的なフォーマットで記述された文書であることを特徴とする、

請求項1記載の解析方法。

【請求項4】 前記電子文書の分割は、当該電子文書がその本文中に同一階層の次の目次項目と同じフレーズ又は定型キーワードが出現するものである場合は、前記出現在するポイントを探し、当該ポイント以降を前記次の目次項目に対応する本文データとして分割する処理であることを特徴とする、

請求項1、2又は3記載の解析方法。

【請求項5】 前記電子文書の分割は、当該電子文書がその本文中に同一階層の次の目次項目と同じフレーズ又は定型キーワードが出現しないものである場合は、前記電子文書の本文を段落毎に分割するとともに、各段落において所定の定型キーワードが存在し且つ当該段落に分割の対象となっている目次項目の名詞句が含まれている割合が1つの前段落の目次項目の名詞句を含む割合よりも大きい場合に当該段落と前記1つ前の段落とを分割する処理であることを特徴とする、

請求項1、2又は3記載の解析方法。

【請求項6】 前記登録の際に、分割された一の本文データに含まれる個々のキーワードと当該キーワードを含む目次項目に対応する電子文書とを相互に関連付けておくことを特徴とする、

請求項1、2又は3記載の解析方法。

【請求項7】 検索対象となる電子文書を保持する電子文書データベースと、

前記電子文書から目次情報を抽出して当該電子文書における目次项目的構造解析を行い、各目次項目に前記構造解析結果に基づく目次レベルを付与する目次情報解析モジュールと、

前記目次レベルに基づいて前記電子文書の本文を目次レベル毎に分割し、分割後の本文データを検索単位として登録する本文分割モジュールとを備えて成る、電子文書登録システム。

【請求項8】 分割された一の本文データに含まれる個々のキーワードと当該キーワードを含む目次項目に対応する

電子文書とを相互に関連付ける関連付けモジュールをさらに備えて成る、

請求項7記載の電子文書登録システム。

【請求項9】 検索対象となる電子文書を保持する電子文書データベースと、前記電子文書に含まれる目次項目の構造を特定するための目次情報解析ルール、前記電子文書における定型キーワードの記述部位に応じた文書分割ポイントを決定するための文書分割ルール、及び、電子文書の分割単位同士の相互関連付けルールを格納したルール格納手段と、

前記電子文書に含まれる目次情報を抽出して目次项目的構造を前記目次情報解析ルールに従って解析し、各目次項目に前記構造解析結果に基づく目次レベルを付与する目次情報解析モジュールと、

前記目次レベルのみ又はその目次レベルと前記文書分割ルールに基づいて前記電子文書の本文を目次レベル毎に分割し、分割後の本文データを検索単位として登録する本文分割モジュールと、

分割された一の本文データに含まれる個々のキーワードと当該キーワードを含む目次項目に対応する電子文書とを前記関連付けルールに従って相互に関連付ける関連付けモジュールとを有し、

前記電子文書を検索前に所定の意味内容単位に分割しておくことを特徴とする、電子文書登録システム。

【請求項10】 前記目次項目が階層的な構造を有するものであり、

前記目次情報解析モジュールは、前記電子文書の目次項目の階層関係を識別するための目次レベルを各目次項目に付与するものであり、

前記本文関連付けモジュールは、前記目次レベルが最下位層であって目次項目と同じフレーズ又は定型キーワードによっては分割ポイントを判定できない場合に前記文書分割ルールに基づいて前記電子文書の本文を分割するものであることを特徴とする、

請求項9記載の電子文書登録システム。

【発明の詳細な説明】

【0001】

【発明の属する技術分野】 本発明は、大容量の文書ファイルを効率的に検索できるようにするための文書登録技術に係り、特に、検索対象となる電子文書の解析方法及び電子文書登録システムに関する。

【0002】

【従来の技術】 日々作成して蓄積された多種多様の電子文書、例えば会議録、報告書、裏議書、論文、マニュアル等を電子化した文書ファイルから利用者の要求するものを検索し、検索した情報を提供する情報検索サービスシステムがある。この種の情報検索サービスシステムでは、全文検索等の手法を用いて該当する情報を利用者に提供するようになっているが、検索結果が、作成時の情報単位、例えば文書ファイル単位で利用者に提供される

場合が多く、サイズが大きい文書ファイルの場合は、真に要求される情報の存在ポイント近くにナビゲートするのが困難であった。このような場合に、例えば利用者の要求度合いに応じて文書ファイルを分割し、検索単位を細かくして登録することも考えられるが、それを実現する場合は、どの基準で分割したら良いかどうかを人間が判断する必要があり、実際に分割する際にも人手を要していた。

#### 【0003】

【発明が解決しようとする課題】上述のように、従来の情報検索サービスシステムでは、検索対象となる文書ファイルが大容量でいくつもの話題を扱った文書、あるいは複数の章を持つ文書であった場合、利用者は、検索結果から真に欲する情報を到達するまでに大量の不要な情報をも確認しなければならなかった。また、検索結果が多すぎて結果に対する2次検索が求められることになると問題があった。一方では、検索結果が大容量の場合、データの転送及びディスプレイ上への表示に大きな負荷がかかっていた。さらに、同一の電子文書から利用者のレベルに応じて複数のファイルを作成する場合、データベース構築のための作業が膨大になり、文書の管理形態が複雑となり、人の手による作業もまた増大するという問題もあった。

【0004】そこで、本発明は、利用者が欲する情報に迅速に到達できるような形態での文書登録を可能にする解析方法及びこの解析方法を応用した電子文書登録システムを提供することを主たる課題とする。

#### 【0005】

【課題を解決するための手段】上記課題を解決する本発明の電子文書の解析方法は、情報検索の対象となる電子文書から目次情報を抽出して当該電子文書における目次項目の構造解析を行い、各目次項目に前記構造解析結果に基づく目次レベルを付与するとともに、当該電子文書の本文を前記目次レベル毎に分割し、分割後の本文データを検索単位として登録することを特徴とする。つまり、事前に電子文書を自動分割して管理しておく、検索時に少ないオペレーションで必要情報を到達できるようになる。

【0006】目次項目が階層的な構造を有する場合、目次レベルは、当該目次項目の階層関係を識別するための識別情報として利用される。この場合において、電子文書がその本文中に同一階層の次の目次項目と同じフレーズ又は定型キーワードが出現するものである場合、電子分割の分割は、定型キーワードが出現するポイントを探し、当該ポイント以降を次の目次項目に対応する本文データとして分割する。一方、電子文書がその本文中に同一階層の次の目次項目と同じフレーズ又は定型キーワードが出現しないものである場合、電子文書の分割は、電子文書の本文を段落毎に分割するとともに、各段落において所定の定型キーワードが存在し且つ当該段落に分割

の対象となっている目次項目の名詞句が含まれている割合が1つ前の段落の目次項目の名詞句を含む割合よりも大きい場合には当該段落と前記1つ前の段落とを分割する。

【0007】検索の便宜を図るため、分割後の電子文書を登録する際に、分割された一の本文データに含まれる個々のキーワードと当該キーワードを含む目次項目に対応する電子文書とを相互に関連付けておくようにしても良い。

【0008】本発明が提供する文書登録システムは、検索対象となる電子文書を保持する電子文書データベースと、前記電子文書から目次情報を抽出して当該電子文書における目次項目の構造解析を行い、各目次項目に前記構造解析結果に基づく目次レベルを付与する目次情報解析モジュールと、前記目次レベルに基づいて前記電子文書の本文を目次レベル毎に分割し、分割後の本文データを検索単位として登録する本文分割モジュールと備え、必要に応じて、分割された一の本文データに含まれる個々のキーワードと当該キーワードを含む目次項目に対応する電子文書とを相互に関連付ける関連付けモジュールを備えて構成されるシステムである。

【0009】電子文書に含まれる目次項目の構造を特定するための目次情報解析ルール、前記電子文書における定型キーワードの記述部位に応じた文書分割ポイントを決定するための文書分割ルール、及び、電子文書の分割単位同士の相互関連付けルールを格納したルール格納手段を備え、これらのルールを使用して上記分割及び関連付けを行うようにすることもできる。

#### 【0010】

【発明の実施の形態】以下、図面を参照して、本発明における実施の形態を詳細に説明する。図1は、本発明の情報検索システムの一実施形態を示す構成図である。この情報検索システム1は、キーボードやポインティングデバイス等の入力装置及びディスプレイパネル等の表示装置を有するコンピュータによって実現されるもので、目次情報解析ルール、文書分割ルール、関連付けルールを格納したルール格納部10、出入力インターフェース11、検索モジュール12、目次解析モジュール13、本文分割モジュール14、関連付けモジュール15、電子文書DB16を備えて構成される。

【0011】出入力インターフェース11は、入力装置や表示装置と本システムとの間で授受されるデータや各種情報の入出力制御を行う。情報検索を要求する利用者端末がネットワーク等を介して接続される場合、出入力インターフェース11は、これらの利用者端末との間で検索条件の入力や検索結果の出力を制御する。

【0012】検索モジュール12は、入力装置を通して利用者から検索要求が入力されたときに、表示装置に検索条件の入力を促すための画面を表示させるとともに、この表示に対応して入力された検索条件に基づく検索結

果（文書ファイル）を表示装置に表示させる。

【0013】目次解析モジュール1.3及び本文分割モジュール1.4は、検索対象となる電子文書（文書ファイル）の自動分割処理を行うものである。目次解析モジュール1.3は、文書ファイルから目次情報を抽出して当該文書ファイルにおける目次項目の構造解析を行い、各目次項目に前記構造解析結果に基づく目次レベルを付与するものである。目次項目の構造解析に際しては、目次項目及びその内容を把握するために文書ファイル内の形態要素毎に品詞を与える公知の形態素解析手法を用いる。目次レベルは、目次項目が構造を特定する文中にどのような形態で出現するかによって定める。例えば、目次項目が本文中の段落先頭に出現する場合は「レベル1」、目次項目に対応す定型語句や記号（以下、「定型キーワード」又は単に「キーワード」と称する）が本文中の段落先頭に出現する場合は「レベル2」、いずれもが該当しない場合は、「レベル3」のように付与する。なお、レベル数は任意であって良い。目次項目が階層的な構造（ツリーコード）を有する場合、目次レベルは、当該目次項目の階層関係を識別するための識別情報として機能する。

【0014】本文分割モジュール1.4は、目次解析モジュール1.3の処理結果及び文書分割ルールに基づいて文書ファイルの本文を分割する。分割方法は、目次レベルによって異なる。これについては後述する。

【0015】関連付け処理部1.5は、関連付けルールに記述されている定型キーワードを含む分割結果とその定型キーワードを含む目次项目的本文情報との相互関連付けを行う。

【0016】次に、上記のように構成される情報検索システム1において、電子文書DB1.6内の文書ファイルを自動分割する場合の動作例を説明する。ここでは、図2に示されるように、文書ファイル2.0に含まれる目次2.1を目次解析モジュール1.3で解析してその目次2.1にそれぞれ目次レベルを付与し、この目次レベルに基づいて、本文分割モジュール1.4で1枚の文書ファイル2.0をM（M：2以上の自然数）枚に分割する場合の例を挙げる。

【0017】図3（a）は目次情報解析ルールの一例、図3（b）は文書分割ルールの一例であり、それぞれ分割対象となる文書ファイルが決定されたときにチューニング（内容の調整・決定）されるようになっている。目次情報解析ルールでは、階層構造のルール、つまりツリー配下をもつ目次のルールも定められている。例えば第1章「ddd」の配下に第1項「ccc」・・・があり、第1項「ccc」の配下に1、「eee」、「fff」がある場合のレベルの決定ルールも定められている。

【0018】文書分割ルールは、主として目次項目がレベル3の場合の分割ポイント（分割部位）を画定する際に用いるルールであり、意味の切れ目である定型キーワードがどの部位に存在した場合に分割ポイントとするかを定めたものである。定型キーワードが文頭にある場合のルールが「文頭ルール」、文中にある場合のルールが「文中ルール」、文末にある場合のルールが「文末ルール」である。文頭ルールの例としては、「次に、それでは」等、文中ルールの例としては、「について、に関して」等、文末ルールの例としては、「何います、お尋ねします」等の定型キーワードが挙げられる。

【0019】目次情報解析モジュール1.3は、図4の手順に従って各目次项目的フラグチェック（目次レベルの付与）を行う。まず、ルール格納部1.0から目次情報解析ルールを読み出す（ステップS101）。そして、目次2.1から1項目ずつ読み出し（ステップS102）、その項目を目次情報解析ルールに従って、3段階のレベルの何れかが、つまりレベル1、レベル2、レベル3の何れかにフラグチェックする（ステップS103）。

【0020】レベル1に該当する場合は、その項目に「レベル1」を付与する（ステップS104a）。レベル2の項目をツリー配下に含む場合はレベル2フラグをONにし（ステップS105a：Yes、S106b）、レベル2の項目をツリー配下に含まない場合は、レベル2フラグをOFFにする（ステップS105a：No、S106a）。

【0021】ステップS103において、レベル2に該当する場合は、その項目に「レベル2」を付与する（ステップS104b）。レベル3項目をツリー配下に含む場合はレベル3フラグをONにし（ステップS105b：Yes、S106c）、レベル3項目をツリー配下に含まない場合は、レベル3フラグをOFFにする（ステップS105b：No、S106d）。ステップS103において、レベル3に該当する場合は、その項目に「レベル3」を設定する（ステップS104c）。以上の処理をすべての項目について繰り返す。

【0022】上記のような目次情報解析結果に基づいて、本文分割モジュール1.4は、目次レベルに応じた文書ファイルの分割を行う。図5は、この場合の本文分割モジュール1.4の処理手順図である。

【0023】本文分割モジュール1.4は、まず、レベル1の分割処理を行う（ステップS201）。すなわち、文書ファイルの本文中（段落の先頭）に次の目次項目と同じフレーズが出現するポイントを探し、当該ポイント以降を次の目次項目に対応する本文データとして分割する。次に、レベル1の分割処理により分割された本文データについて、レベル2または3の目次項目をツリー配下に含むかどうかを判定し、含む場合は、レベル2の分割処理を行う（ステップS202：Yes、S203）。すなわち、本文中（段落の先頭）に次の目次項目を表す定型キーワードが出現するポイントを探し、当該ポイント以降を次の目次項目に対応する本文データとして分割する。

【0024】その後、そのレベル2の分割処理がなされた文書ファイルについて、レベル3の項目をツリー配下に含むかどうかを判定し、含む場合は、レベル3の分割処理を行う（ステップS204：Yes、S205）。

なお、ステップS202においてツリー配下にレベル2又は3の目次項目を含まない場合はレベル1の分割を行い、ステップS204においてツリー配下にレベル3の目次項目を含まない場合はレベル2の分割を行うことになる。

【0025】レベル3の場合は、レベル1及びレベル2の分割のように明確なポイントが存在しないのが通常である。そのため、本文分割モジュール14は、文書分割ルールに従い、図6の手順で分割ポイントを画定する。

まず、ルール格納部10から文書分割ルールを読み出す（ステップS301）。そして、レベル2の分割処理がされた文書ファイルを段落毎に分割し（ステップS302）、各段落において、意味の切れ目となる定型キーワードが前の段落の文末、次の段落の文頭又は文中のいずれかに存在するかどうかを調べる（ステップS303）。

【0026】定型キーワードが上記のいずれかの部位に存在する場合は、当該段落に分割の対象となっている目次項目の形態素である名詞句が含まれているかどうかを調べる（ステップS303：Yes、S304）。名詞句が含まれている場合は目次项目的名詞句を含む割合が、1つ前の目次项目的名詞句を含む割合よりも大きいかどうかを調べる（ステップS304：Yes、S305）。

【0027】なお、目次项目的名詞句を含む割合は、目次项目的形態素解析結果と当該段落の形態素解析結果とを比較することによって求めることができる。例えば、名詞句を対象とし、1つマッチすることによって1点とすることで、割合を比較できるようになる。以上の処理をすべての段落について繰り返し、解析する段落がなくなった時点で分割処理を終了する（ステップS306）。

【0028】次に、関連付けモジュール15において実行される分割結果の関連付け処理の内容について説明する。図7（a）は関連付け処理の概念図、図7（b）は関連付けルールの例を示した図であり、図8は、関連付け処理の手順説明図である。

【0029】関連付けモジュール15は、ルール格納部10から関連付けルールを読み出とともに（ステップS401）、本文分割モジュール14で文書ファイル2から分割された本文データ（レベル3の目次項目に対応するデータ）23～26を順次読み出し（ステップS402）、関連付けルールにマッチする定型キーワードが存在するかどうかを確認する（ステップS403）。存在する場合は、その定型キーワードが他の本文データの目次項目に含まれているかどうかを調べる（ステップS403：Yes、S404）。他の目次項目に含まれて

いた場合は、その本文データと他の目次項目に対応する本文データとを関連付ける（ステップS404：Yes、S405）。この処理をすべての本文データについて繰り返す。

【0030】図7の例では、本文データ23に「XYZの定理」という定型キーワードが含まれており、本文データ25に対応する目次項目に「XYZの定理」という定型キーワードが含まれていた場合に両本文データ23、25の位置情報を相互に関連付ける。なお、「XYZの定理」という定型キーワードは、図7（b）に記述されている関連付けルールである「の定理」の前に記述されている名詞句「XYZ」を連結することによって生成する。

【0031】このように、本実施形態の情報検索システム1では、検索対象となる文書ファイルの目次情報を解析して、事前にその文書ファイルを自動的分割して管理しておくようにしたので、従来のように文書ファイル全体を検索して提示する場合に比べて、利用者の欲しい情報により速く到達することができ、しかもより細かいポイントにナビゲートできるようになる。自動分割に際しては、「巻」、「節」、「章」、「項」、「文」といった、複数の意味的、文法的、形式的固まりからなる大きなサイズの文書ファイルから目次情報を抽出して解析するようになっていたので、人的作業に近い形態で文書ファイルを自動分割することができ（もともと目次情報は、人的作業により当該文書ファイルの意味的、文法的、形式的固まりを判断して作成されたものである）、検索精度も向上する。

【0032】また、文書ファイルを細かく分割しておくことで、検索結果やデータ閲覧のためのファイル転送及び表示の際の負荷を軽減することもできる。

【0033】さらに、分割された本文データの関連付けを行おうにしたので、利用者が目的情報を到達するルートを新たに付加することができるようになる。

【0034】本発明は、例えば会議録、マニュアル、論文、報告書のように、目次項目がほぼ定型的なフォーマットで記述される文書ファイルにおいて、特に顕著な効果を奏すことができる。

### 【0035】

【実施例】次に、本発明の実施例を説明する。上記情報検索システム1は、多様な情報検索に適用することができる。例えば、議員が質問をし、役職者がその質問に回答するという会議の議事録を記録した文書ファイル（会議録）を検索する際に、本システムを応用することができる。会議録は、対象となる文書ファイルが大容量であり、いくつもの話題を扱ったもの（複数の章を持つ文書）であり、目次項目も略定型的なフォーマットで、ツリー構造になっている。会議録としての性質上、質問部分とそれに対応する回答部分とを関連付けることにより、検索の便宜を図る利点も生じる。

【0036】図9は、会議録の自動分割及び関連付けを行う場合に使用される各種ルールの内容例を示したものであり、(a)は、目次情報解析ルール、(b)は文書分割ルール、(c)は関連付けルールである。なお、各ルールは、適宜修正できるようになっている。

【0037】図10(a)は目次解析モジュール1.3によってレベルが付与された目次項目の例である。図示の例では、便宜上、文字の先頭位置によって目次項目のツリー構造を表現している。図中、最左端の目次項目から順にレベル1、レベル2、レベル3が付与されている。

【0038】図10（b）は、図10（a）におけるX議員（レベル2）のソリューションにあるレベル3の目次項目、すなわち、「来年度予算について」a1と「行政改革について」a2の本文データの例を示した図である。この図10（b）の例について、どのようにして両者の本文データを分割するか、つまりレベル3分割を行うについて説明する。

【0039】分割対象となる図10（b）下段の本文データにおいて、「次に」b1は文頭ルールにマッチし、「に関連して」b2は文中ルールにマッチする。一方、上段の本文データにおける「お問い合わせ」b3は文末ルールにマッチする。

【0040】また、下段の本文データには、次目項目の名詞句である「行政改革」**b**が含まれている。この「行政改革」が含まれる割合は、上段の本文データに対応する次目項目である「来年度予算について」よりも大きい。つまり、図6に示したステップS303～S305の3つの条件をすべて満足している分割ポイントPを境として分割を行なう。

【0041】次に、関連付け処理の例を説明する。図11(a)は次目解釈モジュール1.3によってレベルが付与された次目項目の例である。最左端の次目項目から順にレベル1、レベル2、レベル3が付与されている。図11(b)は、レベル3の次目項目「震災対策について」を1.1の本データ、つまりXX議員の質問データーの一部を記述したものである。文面中の「Y局」はb-1.1は、関連付けルールに記述されている定型キーワードであり、図11(a)の他の次目項目a12にもその定型キーワード「局長」が含まれているので、この次目項目a12に張り付けられている分割結果と互いに位置情報を保持し合う。つまり、2つの分割結果は、条件を

満足することで関連付けられる。

(0042)

【発明の効果】以上の説明から明らかのように、本発明によれば、大容量の電子文書が事前に自動分割され、しかも関連性のあるものについては相互に関連付けされているので、必要とする情報を少ないオペレーションで索出して利用者に提示できるようになる。

### 【図面の簡単な説明】

【図1】本発明の一実施形態に係る情報検索システムの構成図。

【図2】本実施形態による自動分割処理の概念図。

【図3】(a)は目次情報解析ルールの

文書分割ルールの一例を示した説明図。

【図4】本実施形態による自動分割処理の手順説明図。

【図5】本文分割モジュールによる分割処理の手順説明

图。

【図6】本実施形態によるレベル3の分割処理の手順説

明図。

【図7】(a)は関連付け処理の概念図、(b)は関連

付けルールの一例を示した説明図。

【図8】本実施形態による関連付けの処理の手順説明

图。

【図9】(a) 本発明の実施例で用いる目次情報解析ル

ール、(b)は文書分割ルール、(c)は関連付けル

ルの一例を示した説明図。

【図10】(a)はレベルが付与された目次項目の例、

(b)は、XX議員(レベル2)のツリー配下にあるレ

ベル3の二つの目次項目の本文データの例を示した図。

【図11】(a) はレベルが付与された目次項目の例、(b) は各項目の子項目を示す。図11(b) は、(a) の子項目を示す。

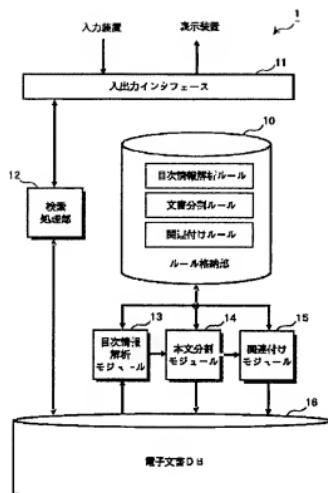
(b) は、レ

- 本文データの例を示した図。

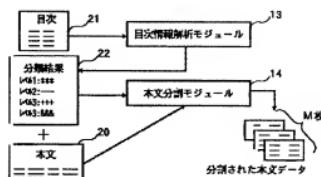
【符号の説明】

  - 1 情報検索システム
  - 10 ルール格納部
  - 11 入出力インターフェース
  - 12 検索モジュール
  - 13 目次情報解析モジュール
  - 14 本文分割モジュール
  - 15 関連付けモジュール
  - 16 電子文書DB（データベース）

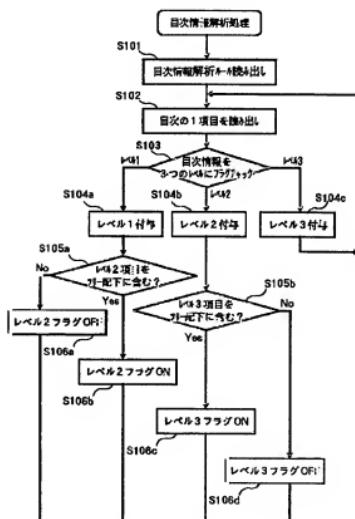
【図1】



【図2】



【図4】



【図3】

(a)

```

<レベル3の目次を> 配下に持つ目次>
aaa
bbb
.....
<レベル2の目次>
ccc
.....
<レベル2と3の目次を> 配下に持つ目次>
ddd
.....
<レベル3の目次>
eee
fff
.....

```

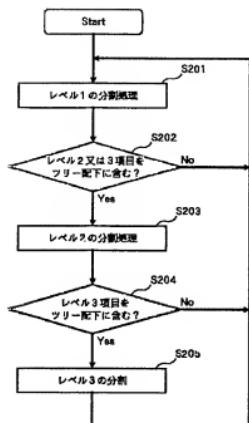
(b)

```

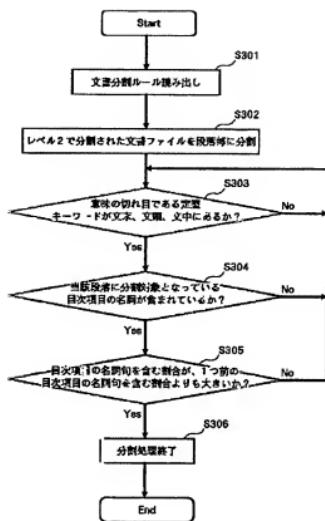
<文頭ルール>
ooo
hhh
.....
<文中ルール>
ii
.....
<文末ルール>
jj
.....

```

【図5】



【図6】



【図9】

(a)

<レベル1項目> 配下持つ項目  
 “西郷の上り”  
 “道具登場”  
 “夏至音頭歌”  
 “火祭”  
 “射撃”  
 “百合のあいさつ”  
 “旅館の泣き上り”  
  
 <レベル2項目> 配下持つ項目  
 “實業社説”  
 “政治”  
  
 <レベル2の次>  
 “\*”(定期刊号)、“雑誌”、“知事”  
 “財政”、“征税”、“船員”、“局長”  
 “委員会”、“教育長”、“地主”  
  
 <レベル3の項目>  
 “1”、“2”、“3”、“4”、“5”、“6”、“7”、“8”、“9”

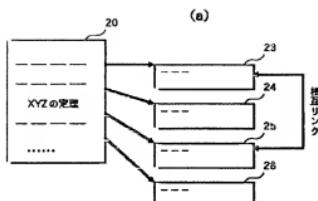
(b)

<文頭ルール>  
 “火に。”  
 “さて次に。”  
 “さて次は。”  
 “最後に。”  
  
 <文末ルール>  
 “御所見を。”  
 “お尋ね”  
 “お問い合わせ”  
 “問って”  
 “お聞き”  
 “答弁”  
 “お聞かせ”  
  
 <文中ルール>  
 “ついで”  
 “つきまして”  
 “お聞達して”

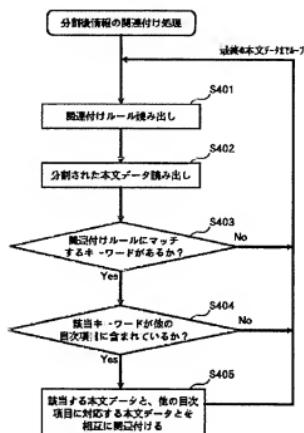
(c)

“知事”  
 “出知事”  
 “本部長”  
 “委員長”  
 “支局長”  
 “教育長”  
 “船員”  
 “局長”  
 “旅館”

【図7】

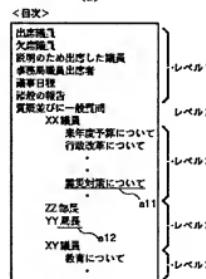


【図8】



【図11】

(a)

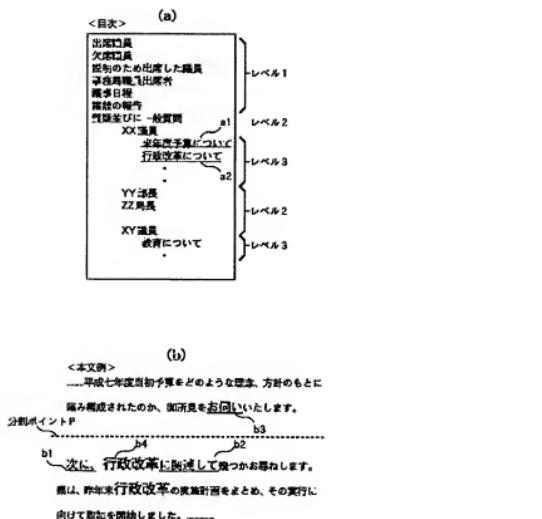


(b)

<本文例>

|  |
|--|
| XX地図<br>【1. 避災対策について】<br>—YY地図の答弁をお詫び致します。<br>—b11 |
|--|

【図10】




---

フロントページの続き

(72)発明者 小室 英一

東京都江東区豊洲三丁目3番3号 株式会  
社エヌ・ティ・ティ・データ内

(72)発明者 山中 航史

東京都江東区豊洲三丁目3番3号 株式会  
社エヌ・ティ・ティ・データ内

Fターム(参考) 5B009 ND01 QA09 QA12 SA12 VA02

5B075 ND03 NR02 NR12 PR10